

How does EVAAS accommodate changes in testing regimes?

Introduction

EVAAS value-added analyses have been available to educators and policymakers for the past two decades. During that time, there have been many changes in testing regimes in the states that use EVAAS. From this experience, SAS has developed a statistically robust process to utilize all available testing data from both old and new tests as well as to ensure continuity in reporting. This document provides an overview of how EVAAS accommodates changes in testing regimes, but please note that the actual EVAAS models are more statistically robust than the simplified examples represented here.

How do EVAAS models typically calculate value-added?

EVAAS value-added measures are calculated using all available student test scores across all grades and subjects for the most recent five years. The multivariate response model (MRM) is a gain-based approach, and it is typically used for consecutive grade-given tests like math and reading in grades three through eight. To measure gains from one grade to the next, the student scale scores are first converted to Normal Curve Equivalent (NCE) distribution based on the state population for a particular test/subject/grade/year. The NCE distribution matches a percentile distribution at 1, 50 and 99. However, unlike a percentile distribution, the NCE distribution is equal-interval such the difference between 50 and 60 is the same as the difference between 80 and 90. This is an important distinction for measuring gains to ensure that educators are neither advantaged nor disadvantaged by the entering achievement of their students. See Figure 1 for a graphical representation of these distributions.

To simplify the MRM approach, the average NCE gain for students with a district, school or teacher is based on their change in achievement, as measured in NCEs. For example, if students in a school entered at the 50th NCE and left at the 55th NCE, then the value-added measure would be the difference between 55 and 50, or 5 NCEs. If students entered at the 50th NCE and left at the 50th NCE, then their gain would be zero. However, this does not represent zero growth; it simply means that students have maintained their position in the distribution. Typically, zero represents the growth standard or expectation for students.

However, setting the growth expectation is a bit more complicated. The NCE distribution itself can be based on either a “base year” or an “intra-year” approach, and each is described briefly below.

- Base year: a cohort of students from grade to grade maintains the same relative position with respect to the statewide student achievement in the base year for a specific subject and grade.
- Intra-year: students maintained the same relative position with respect to the statewide student achievement that year.

If a base year approach is used, then the student scale scores are translated into NCEs based on the distribution of scores from the base year, and the gains are thus relative to the achievement in the different grades of the base year distribution. Because the tests are given in the same scales each year, it is possible to measure gains relative to the base year.

If an intra-year approach is used, then the base-year is essentially re-set each year so that gains are considered in terms of their relative position each year. Because the gains are relative, it is not necessary for the tests to be in the same scale in measuring gains from one year to the next.

Again, in either approach, an average NCE gain is calculated for a district, school or teacher in a given test/subject/grade/year as a value-added measure. This value is compared to a growth standard of zero,

which represents that this group of students, on average, maintained the same relative position in the statewide distribution from one grade to the next in a given subject.

The MRM approach is typically used for tests given in consecutive grades. For tests that are given in non-consecutive grades, such as end-of-course exams or non-consecutive end-of-grade exams, the univariate response model (URM) can be used to measure growth. Rather than a gain-based model, the URM is a regression-based model where the growth measure is based on the difference between students' predicted achievement and their observed achievement. The predicted achievement is based on students' prior testing history, and these do not have to be on the same scale. The expectation of progress is based on how students in the reference population did in the current year. In this way, the URM can be thought of as somewhat similar to the intra-year approach. In other words, there is not a comparison to a base year; the growth expectation is determined within the current year. The growth expectation of the URM can be described as follows:

- Intra-year: students in a district/school/teacher made the same amount of progress as students in the average district/school/teacher in the state for that same year/subject/grade.

In URM, as in MRM, a value-added measure of zero does not indicate that students made no growth. In URM, a value-added measure of zero indicates that students made the expected progress or similar progress as students with the average district, school or teacher (depending on the model). URM does not use NCEs, and its value-added measure is always reported in the scaling units of the current test.

How do EVAAS models typically calculate value-added in a transition year?

In a transition year, EVAAS will continue to use all available student test scores across all grades and subjects for the most recent five years, and the MRM (or gain-based) approach can still be used. However, because the historical data are not on the same scale as the current year's test scores, it is not possible to use a base year approach during the transition year. However, the intra-year approach can still be used because the NCE distributions for each year are based on that specific year. For example, if measuring gains from 2012 to 2013, the scores from 2012 are used to create a 2012 NCE distribution and the scores from 2013 are used to create a 2013 NCE distribution. The value-added measure for 2013 is based on the change in achievement from the 2012 NCE distribution to the 2013 NCE distribution. The intra-year approach takes the scaling units of the test out of the modeling.

The intra-year approach is typically recommended for the first two years after a new test is implemented. If test scores appear stabilized after the second year of implementation, then the base year approach can be used again.

Because the URM currently uses only an intra-year approach and reports value-added estimates in the scaling units of the current test, a change in testing regimes does not affect the modeling or reporting. The URM reporting can continue as always during the transition year.

Does the release of new standards affect value-added reporting?

No, the release of new standards or performance levels does not affect value-added for students. Value-added measures are calculated from actual student scores; they make no use of the performance levels.

Figure 1: Sample Distribution of Scores, NCEs and Percentiles

